2020

Non-destructive, Rapid Differentiation of Cell Types Relevant to Sexual Assault Investigations Utilizing Morphological and Autofluorescence Signatures

Taylor R. Moldenhauer
Virginia Commonwealth University

Non-destructive, Rapid Differentiation of Cell Types Relevant to Sexual Assault Investigations Utilizing Morphological and Autofluorescence Signatures

Taylor Moldenhauer
Summer 2020-Fall 2020
November 20th, 2020

Mentor: Christopher J. Ehrhardt, Ph.D.
Ehrhardt Research Laboratory
Department of Forensic Science
Virginia Commonwealth University

A thesis/dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in Forensic Science at Virginia Commonwealth University.

Acknowledgements

I would like to thank Dr. Christopher Ehrhardt for his endless help and encouragement as well as for inspiring me to think more broadly in everything I do. Additionally, I would like to thank my committee members, Kate Philpott, JD, Dr. Catherine Connon, and Dr. Susan Greenspoon, for all their assistance throughout this project. Thank you to the Flow Core—Julie Farnsworth and Dr. XinYan Pei—for their patience, help, and conversation as I processed my samples. My sincerest appreciation to Justin Mehojah and Kayla Burns for challenging me to grow and learn with every new obstacle and opportunity. Lastly, my never-ending gratitude to my mother Tiffany for being my biggest fan, my best critic, and the reason for my success. Thank you for enabling me to find my own path with nothing but support and understanding.

Abstract

Current methods for confirming the presence of spermatozoa in sexual assault samples can be time-consuming and often lack sensitivity; however, this remains the most definitive test for the presence of semen. Additionally, male DNA can be deposited without the presence of intact sperm as may be the case with seminal fluid from vasectomized individuals or sexual activity where seminal fluid is not recovered (e.g., perpetrator wears a condom, penetration without ejaculation, etc.). The ability to detect bodily fluids, as well as quantify their presence in a sample, could aid in forensic DNA analysis by limiting the amount of time performing serological testing, as well as screening for probative samples for DNA profiling. Additionally, determining the cellular makeup of a sample can be informative for investigative purposes, e.g. rebutting or supporting certain factual claims from the victim or defendant. Morphological and/or autofluorescence cellular signatures are rapid and non-destructive methods for cell type differentiation in the clinical context but have not been thoroughly explored for forensic casework applications. Therefore, the goal of this study was to characterize signatures in four major cell types associated with sexual assault casework (vaginal, rectal, and penile epidermal cells, and azoospermic seminal fluid) towards the development of a method for rapidly identifying and/or differentiating these cell/fluid types in biological evidence. Morphological and autofluorescence profiles of each cell population were analyzed with Imaging Flow Cytometry (IFC) using five different excitation wavelengths and six detector channels ranging between 430nm-780nm. Signatures for each cell type were constructed from ~60 different individual cell measurements. Finally, linear discriminant analysis was used to develop a quantitative framework for differentiating cell populations and predicting cell types. Vaginal, rectal, and penile cells can be differentiated with a high degree of accuracy, ~90%. This framework was also highly accurate at differentiation semen (including azoospermic and proteinase K treated semen) from vaginal and rectal cell populations. However, there were still many factors that contributed to these levels of accuracy including, but not limited to, inter-donor variability. Ultimately, the results obtained indicate that each cell type have distinctive signatures that can be detected in a rapid and non-destructive manner.

Key Words: cell differentiation, forensic serology, sexual assault, sperm, flow cytometry

Introduction

In sexual assault cases the presence of sperm is an indicator of sexual activity and, ultimately, the presence of male DNA. With a large backlog of sexual assault kits still existing in many states, the development of new methods that can expedite sample processing is an ongoing priority for the forensic biology community [1, 2]. Additionally, there are many case working scenarios where sexual assault has occurred, but the perpetrator's semen is not deposited within evidentiary samples, e.g., sexual contact without ejaculation, azoospermic semen deposit. In these instances, evidence samples may be mixtures of epithelial cells deposited by the perpetrator (e.g., epidermal layers derived from the hands and/or penis) and epithelial cells from the victim (e.g., vaginal swab, rectal swab). Although there are many strategies for analyzing sexual assault samples, most have significant limitations with respect to the speed, sensitivity, and/or the range of samples that can be characterized. Some of the most common methods are summarized below.

Serological Methods

Traditional serological testing involves a sequence of presumptive tests and confirmatory tests followed up with DNA analysis if the target fluid's presence has been confirmed. However, a common trend for forensic laboratories is to proceed from a positive presumptive test directly to DNA analysis—forgoing all confirmatory testing for efficiency. This means a great deal of reliance must be placed on these tests. The most common presumptive test for semen is the Acid Phosphatase test (AP), which has a high sensitivity but a low specificity as there are chances for false positives including detection of acid phosphatase in other fluids (vaginal fluid and urine) that have a possibility of being present for sexual assault casework [3, 4] It has been shown that even when spermatozoa are positively identified through microscopic examination, the AP test can be negative [5, 6]. Additionally, as samples age the acid phosphatase present degrades, which can

cause the AP test to be inconclusive or negative—especially for cold case samples [7]. Other presumptive tests are far less accurate and typically precede an AP test. Utilizing alternate light sources (ALS) or UV light to identify stains on samples can see an array of substances fluorescing: saliva, semen, yogurt, and milk are just a few of the kinds of stains that all "glow" under UV light [8].

When labs do choose to conduct confirmatory testing, it can be either a microscopic examination to visualize sperm or an antibody-based test for the presence of Prostate Specific Antigen (PSA or p30) such as ABAcard® p30, Seratec® PSA Semiquant, and RSID-Semen [7, 9, 10]. These methods can more accurately note the presence of sperm/seminal fluid—definitively indicating male biological material is present. However, the biggest drawback is that they are time consuming (e.g., sample preparation alone can take over an hour). Another challenge with processing sexual assault evidence occurs when male biological material is deposited without the presence of intact sperm as may be the case with seminal fluid from vasectomized individuals or sexual activity where seminal fluid is not recovered (e.g., perpetrator wears a condom, digital penetration, etc.). These samples can cause a considerable amount of time to be spent searching for sperm where none may be found [9]. While a p30 card could then be utilized to confirm the presence of seminal fluid, it means additional sample preparation and consumption [11]. Additionally, cases where male DNA is present due to the presence of shed penile skin cells rather than semen would likely provide negative results for the above presumptive and confirmatory tests.

Y-STR Analysis

One recent strategy to overcome some of these challenges is Y-STR screening. In cases of low sperm count, vasectomized ejaculation, penetration without ejaculation, condom usage, or aged samples, male DNA detection can be hindered by and little to no sperm cells and an

overwhelming presence of female DNA [12]. Y-STR analysis is a sensitive method of analysis, only requiring a small amount of DNA to achieve a full profile, making it an invaluable tool for sexual assault casework. Y-STR analysis' use in sexual assault cases is typically only considered after the autosomal STR profile is obtained and a low level mixture is noted where there is concern a male profile is still present but not represented [12-14]. Therefore, having a method for identifying and quantifying azoospermic, low count, or degraded samples rapidly could prove beneficial when assessing the DNA analysis needs of samples in a case.

Direct to DNA

Direct to DNA is another recent approach recommended by SWGDAM for rapid processing of sexual assault kits. Here, samples go straight to DNA extraction and profiling without presumptive serological testing [15, 16]. The aim of Direct to DNA is to obtain CODIS-searchable profiles from as many samples as possible. While this is a step forward for rapidly testing sexual assault cases, traditional serological testing (e.g., AP test) may still be performed on a cutting of the original swab to provide probative context for DNA profiles after they are obtained.

Molecular RNA markers

Some current research into differentiating bodily fluids based on molecular markers has been focused on utilizing RNA-based markers including messenger and micro RNAs [17-19]. Messenger RNA targets have shown promise in detection of tissues and cell types, including distinguishing vaginal cells from buccal cells [20]. The biggest drawback to mRNA utilization, however, is the cross-reactivity of markers observed across different bodily fluids [21]. Micro RNAs are beginning to be explored for various types of tissue differentiation [22]. Like mRNAs, microRNAs are not completely tissue specific; further, microRNA experiments have

reproducibility issues [23]. Micro and mRNAs have shown great promise for detecting the presence of semen including infertile semen [20, 21]. RNA can be extracted concurrently with DNA utilizing the same tube and kits [17, 22]. It does, however, involve additional extraction and cleanup steps as well as an additional RT-PCR step [24, 25]. Additionally, long term stability of RNA is not noted as it is for DNA, leading researchers to consider DNA-based molecular methods for distinguishing fluids and tissues [23, 26].

DNA methylation

Another molecular method for tissue differentiation includes DNA methylation markers. DNA methylation contributes to cellular differentiation and therefore, looking at the methylation patterns found in different fluids and cell types can be useful in distinguishing sample types [17, 23, 26]. For example, methylation patterns in skin cells are distinguishable from those in vaginal and buccal epithelia. Additionally, determining age of donor can be achieved through an assessment of DNA methylation levels [27, 28]. However, it does have its limitations as factors such as age of the sample and environment factors have been noted to affect levels of methylation. Ultimately, much like mRNA and micro RNA, additional steps must be taken by an analyst—either performing High Resolution Melt qPCR or pyrosequencing—post DNA extraction and purification [29, 30].

Imaging Flow Cytometry

A promising, but unexplored method for rapidly characterizing cell types recovered from sexual assault kit samples involves profiling of autofluorescent and morphological properties using Imaging Flow Cytometry (IFC). In IFC, individual cells are interrogated by excitation lasers, causing some to fluoresce at specific wavelengths. Microscopic images are taken of each cell in

each fluorescent channel as well as a brightfield channel [31]. As with other flow cytometry methods, IFC can be performed in a high throughput and non-destructive manner.

Previous studies have demonstrated that IFC can be used to capture a series of morphological and autofluorescence measurements from dried/aged biological samples that permit differentiation of epithelial cell populations from different tissues (buccal, vaginal, touch epidermal) [32]. Although this approach has not been explicitly tested on non-sperm cell populations relevant to sexual assault casework (e.g., vaginal, rectal, penile epidermal), there are several morphological and histological differences across these cell types. For example, vaginal and rectal mucosa are categorized as stratified squamous epithelium alongside the outer layers of epidermis. However, they are distinguished from outer epidermal layers by lesser degree of keratinization [33-36]. Something as simple as cell size can lead to quick distinction as well—previous IFC interrogation of cells showed differences in size for vaginal/buccal compared to epidermal cells. Rectal epithelial cells have been found to often be smaller than buccal, vaginal, and most epidermal cells. Additionally, biochemical distinctions can be made between vaginal/rectal populations and penile or other epidermal cell populations based on glycogenation levels—as this is common to non-keratinized squamous cells [37]. The detection of glycogenated cells was once considered a possible presumptive test for vaginal intercourse when penile swabs were evaluated, however, there are varying levels of glycogenation in oral, anal, and male urethra mucosal cells. However, utilizing/understanding how biochemical compounds fluoresce and accumulate on different cell types may lead to better differentiation [37-39].

As for seminal fluid there are a few cell types present aside from spermatozoa. Cells shedding from the urethra contains a mix of columnar and transitional epithelium. Additionally, previous studies have made reference to other cells known as “round cells”—where the abundance

and specific type are determined by many factors such as age, fertility, and health of the individual [40]. These round cells appear to be common in seminal fluid (including vasectomized) and are most likely lymphocytes, macrophages, neutrophils, or immature sperm cells. Their presence may also be of value as a Y-STR profiles can be obtained from these cells [40, 41] and are potential targets for morphological autofluorescence-based signatures.

The goal of this study was to characterize differences in morphological and autofluorescence profiles from cell populations derived from semen as well as vaginal, rectal, and penile tissue sources using IFC. A multivariate statistical approach was then used to detect signatures across each cell type and test a classification framework for predicting cell types within simulated mixtures. As a final experiment, this approach was also used to characterize differences in cell populations derived from azoospermic semen stains and epithelial cell populations.

Methods

Sample preparation

Samples for this study were provided by the Ontario Centre for Forensic Science from 12 volunteers (six male and six female). A total of five semen swabs (two classified as azoospermic) and six each of penile, vaginal, and rectal swabs were collected from the participants. All samples were swabbed and dried then stored at room temperature (approximately 2 weeks). Finally, samples were packaged and transported and remained in the lab until analysis. Additionally, three Proteinase K treated semen samples (supplied by the previous donors) were treated with Proteinase K to simulate degradation.

To elute the cells from swabs, 1.5 mL of 1x PBS was added to a 50 mL conical tube and the swab, cotton side down, was inserted into the buffer. The samples were then vortexed for approximately 30 seconds, moving the swab around the tube to further agitate. The swab was removed, and the supernatant moved into a 2 mL Eppendorf tube and centrifuged at 13,000 rpm at 4°C for five minutes. The supernatant was removed, taking care to not disturb the pellet, and 500 µL of fresh 1x PBS added. The pellet was resuspended and then centrifuged again as before. The supernatant was removed, and the pellet resuspended in 200 µL of 1x PBS.

Sample analysis and statistics

Samples were analyzed on an Amnis® Imagestream X Mark II (EMD Millipore; Burlington, MA) Imaging Flow Cytometer (IFC). This instrument has 405nm, 488nm, 561nm, and 642nm lasers with voltages set to 120mW, 100mW, 100mW, and 150mW. Six total channels were used to capture images at 40x magnification: channel 1 (430-505 nm), 2 (505-560nm), 3 (560-595nm), 5 (640-745nm), 6 (745-780nm) and channel 4 for Brightfield images. The raw image (.rif) files are imported into the IDEAS® Software (EMD Millipore; Burlington, MA). For all the images, in all channels, the Image Gallery Properties had Image Display Mapping fixed through 'Set Range to Pixel Data'. A scatterplot was created using Area_M04 x Aspect Ratio_M04 to assess the collected image population for a sample. Given that data from IFC can include all particles within a broad size range (~1µm to ~200µm) within a cell population, we first identified a 'gate' or size region in which the target cell type were most likely found and can be differentiated from either larger particles (e.g., epithelial cells) or smaller ones (bacteria, cell debris). Two gates were created for the following samples: a small cell gate (i.e., sperm and semen cell types, cellular debris) and a large cell gate (i.e., vaginal, penile, and rectal epithelial). An example of the population of cell 'events' imaged from a semen sample in the small cell gate is shown in Figure

1. Finally, only the focused cell images that fell within the gated population were selected by creating a Gradient RMS_M04Ch04 histogram for the scatterplot. The number of analyzable cell images varied among samples and/or cell types and are shown in Table 1.

The IDEAS® Software was utilized to extract specific cell measurements for each individual cell event. In all, there were over 20 different feature values acquired for individual cell events including area, diameter, and intensity (See Table 2 for full list of exported feature values). Since most feature values were captured on multiple detector channels, this yielded > 50 measurements for each cell. Once measurement values had been exported, multivariate statistical analysis of cellular measurements was performed to identify signatures for differentiating cell populations (e.g., Discriminant Function Analysis) utilizing SPSS v23 (IBM, Inc. Chicago, IL). Different combinations of feature values were utilized to determine the most accurate classification and greatest differentiation of all cell types.

To complement the differentiation observed and assess whether signatures could be used to predict cell type in unknown samples, we generated predictive algorithms were generated that assigned samples into one of the cell/fluid type groups (i.e., semen, vaginal, rectal). The accuracy of the algorithms was assessed by cross-validated classification tests whereby individual samples are removed from the analysis and then tested as unknown samples (i.e., 'jack-knifed', 'leave-one-out' analysis).

Results

I. Differentiation of semen from vaginal and rectal cell populations

Sperm cell images varied in the presence of a tail (which can reach lengths of 50 µm). However, observed head size was approximately 5 µm. To identify which events in the area-aspect ratio plot corresponded to sperm, brightfield images were visually inspected for the presence of distinct head with tail. Clusters of sperm were observed more frequently than individual sperm images. Additionally, the presence of cells with a diameter size of ~10-15 µm were observed which is consistent with epithelial cell or immune cells that have been found in seminal fluid [14].

Vaginal and rectal cell populations from this gate were mostly small cells and cell fragments. Vaginal cell populations had much more microflora (e.g., bacteria, yeast)/debris captured with the cell events. Rectal cell populations, on the other hand, had clear small cell fragments as well as many more clusters of debris. There were no nuclei noted for either populations. As for fluorescence differences, vaginal cells (within the small cell gate) show strong fluorescence properties in channels 02, 03, and 05. Rectal cells (within the small cell gate) have fluorescence in channels 02, 03, 05. In addition, channel 06 also showing fluorescence properties. Occasionally, rectal cell population events were brightly fluorescent across all five fluorescence channels. The rectal samples overall had far more intense fluorescence noted than vaginal. Semen, on the other hand, had generally low levels of fluorescence in all channels, with occasional higher levels of fluorescence observed in channel 03. Upon examination of cell images, this appeared to originate from sperm cell heads (some imaged singularly and some as a cluster of heads) and by the presence of individual epithelial cells (data not shown).

To increase the relevance for forensic casework, semen samples were also prepared by treating with Proteinase K prior to analysis. Proteinase K has been used to simulate degradation that occurs to sperm in the vaginal cavity prior to sample collection [42, 43]. Images of these samples showed sperm heads (~5 µm), and multi-cell aggregation. However, sperm tails were not

observed, consistent with observations of casework samples [5]. Median fluorescence intensities in all channels were even lower than untreated samples (e.g., 'Intensity', 'Brightness Detail Intensity') which may be due to partial or complete breakdown of endogenous fluorophores within the sperm cells [44].

Next, multivariate analysis of morphological and autofluorescence measurements was conducted to characterize and maximize differentiation of these three cell types. Discriminant Function Analysis (DFA) for untreated semen, vaginal, and rectal populations showed statistically significant differences in multivariate means (Wilk's lambda = 0.316; $p < 0.0001$), with the largest difference between semen and vaginal cells. A plot of all three sample groups using the first two discriminant functions shows distinct clustering with some overlapping distributions of samples among each of the groups (Figure 2). The vaginal population centroid has some separation from the other two centroids and a distinct population cluster that was noticeably separate from the other two overlapped clusters.

To determine whether the differences could be used to build a predictive framework for classifying unknown samples into one of the three sample groups, classification discriminant functions were analyzed. Individual cell events within semen cell populations were correctly identified against vaginal and rectal samples with an accuracy of ~94% (Figure 2). Vaginal and rectal cells had lower classification accuracies of ~76% and 72% respectively.

Since some of the above misclassifications were exclusively between rectal and vaginal cells, two-group classifications were also tested for semen differentiation against vaginal or rectal cells exclusively (Figure 3). For semen-vaginal cell classification, results showed a moderate increase in classification with 98% for semen and 84% for vaginal cells. Similarly, for semen-

rectal classification, classification accuracy for semen was 97% and rectal cells was 78.7%, respectively (Figure 4).

Comparisons between proteinase K treated semen versus vaginal and rectal cells showed stronger multivariate differentiation (Figure 5). Differences in group centroids were significant (Wilks lambda = 0.062, $p < 0.0001$). There is clear clustering of the semen population with some overlap over the rectal population, and the centroid is further removed as well. Vaginal and rectal cell populations have the appearance of major event overlap, and their centroids are relatively close. However, the vaginal instances have a clear, separate group from their main cluster and centroid.

Proteinase K treated semen samples were compared against vaginal and rectal cell populations and initially showed comparable accuracies to the comparisons with untreated semen samples: semen ~97%, vaginal 80%, rectal 70% for individual cell events within each population (Figure 5). However, many of the misclassifications were vaginal cells identifying as rectal cells and vice-versa. Two group classifications showed markedly higher accuracies. When proteinase K treated semen samples were compared against solely vaginal samples the accuracy of classification was ~99% for both cell types (Figure 6). Additionally, proteinase K treated semen samples and rectal cells both had a classification accuracy of 99% (Figure 7).

An additional experiment compared azoospermic samples with vaginal and rectal cell populations to determine if differentiation could be accomplished. Azoospermic samples showed cellular debris and a high population of round cells and epithelia. Cells resembling intact sperm were not observed. Autofluorescence for these samples was also notable across all channels; with some images having bright, obvious fluorescence, which was predominantly observed in channels 02, 03, and 05. This finding corroborates previous work that detected semen autofluorescence

emissions occurring around 530 and 622 nm due to components such as flavins [45]. When compared to the small cell gated vaginal and rectal cell populations, imaged cells for the azoospermic samples were larger and less cellular debris was imaged.

The discriminant function analysis between all three cell types shows distinct separation of the azoospermic cell populations from the vaginal and rectal cell populations. Although not visualized, the vaginal centroid indicates a sizeable number of events overlapping with the rectal cell events. Statistically significant differences were observed between the multivariate means of three group centroids (Wilk's lambda = 0.249; $p < 0.0001$). When the three sample types were assessed together, both vaginal and rectal populations had a classification accuracy between 70% and 77% while the azoospermic samples saw a better classification at 92.8% (Figure 8). However, again, when azoospermic was assessed with vaginal or rectal samples, the classification accuracies were far improved. When azoospermic semen samples were assessed against vaginal samples, the accuracy of classification was 96% for azoospermic semen and 98% for vaginal cells (Figure 9). Additionally, rectal cells assessed with azoospermic semen had an overall classification accuracy of 96% for azoospermic semen and 98% for rectal cells (Figure 10).

II. Differentiation of penile epidermal cells from vaginal, rectal cell populations

For the second experiment epithelial cells from vaginal, penile epidermal, and rectal tissue were characterized. The large cell gate was developed to select epithelial cells greater than 20 µm in size to capture most of the cellular population for these sample types (Figure 11). Similar to previous results, each cell population had distinct attributes that were visible in the brightfield images [31]. Vaginal cells were typically 40-50 µm and rectal cells were > 20 µm, consistent with published literature [46, 47]. Penile epithelia are noted to be around 30-40 µm, however, the

images and further measurements showed a population of cells that were smaller than vaginal and rectal epithelia [46]. This could possibly be due to the limited cell population obtained from the swabs. Nuclei were nearly always present for vaginal cells—with some apparent folding present (Figure 12). This is different from the vaginal small cell fraction that were not observed with nuclei, likely due to this population being made up of cell fragments and debris. Rectal and penile epithelium were rarely found with visual evidence of nuclei.

Individual measurement comparisons showed distinct differences in area and diameter measurements as well as differences in 'Bright Detail Intensity R3' and 'Spot Intensity Max' across fluorescence channels. This indicates variation in their fluorescent profiles—especially seen in channels 02, 03, and 06. Lower quantities of penile cells were infrequently were observed, possibly be due to low shedding of penile epidermis or perhaps has a correlation to donor-specific attributes, consistent with previous studies due to the correlation with low DNA yield [46]. Penile and rectal samples were also noted to have higher quantities of smaller cell events than vaginal cell populations which could be due to the contribution of cellular, microbiota, and/or miscellaneous debris. Issues with vaginal samples on a donor-to-donor basis were also noted—discrepancies in cell count, proportion of smaller cell events,—that may have contributed to poor statistical analysis and were an obvious departure from other vaginal swabs.

Multivariate differentiation shows clear separation between the centroids for all three cell populations (Figure 13). While the vaginal cells appear easily distinguishable from rectal and penile cells, some of the vaginal cell events showed overlapping distributions the penile and rectal populations. The same can be said for rectal cell events overlapping with penile cell events. Regardless, comparisons of multivariate means indicate statistically significant differences between all three group centroids (Wilk's lambda = 0.093; $p < 0.0001$).

Individual cell events within rectal, vaginal, and penile epidermal populations had an overall classification accuracy of 89% (Figure 13). As was seen from the combined group semen experiments, vaginal and rectal cell populations show clusters with more overlap compared to other cell populations. Considering the proximity of vaginal/rectal tissue on the body and that they are both mucosal epithelia, vaginal and rectal cells have many histological and/or biochemical similarities that are difficult to resolve using IFC signatures. Penile cells were typically successful in their separation (84% accuracy). Further two-group comparisons were conducted for vaginal and penile samples as well as penile and rectal samples (Figures 14 and 15). For penile and rectal cells, the classification accuracies went up significantly: penile cells 94% and rectal cells 95%. Similar results were observed for vaginal and penile sample comparison: 96% for vaginal cell events and 94% for penile epidermal cell events.

Other DFA and Classifications

Additional analyses whereby entire donor cell populations were analyzed in a blinded fashion were conducted to characterize the extent of donor-to-donor variation in autofluorescence and morphological signatures. Donor-to-donor heterogeneity has been observed within tissue-type before in IFC studies [31]. The first test of the predictive algorithms was conducted with the proteinase K treated semen and vaginal samples. Cell events within each of the blinded donor cell populations were identified with an accuracy of >98% indicating low inter-donor variability amongst these semen samples (Table 3). This may indicate consistency in signatures across donors, however, a larger test set would need to be obtained to confirm this. The second test conducted was with blinded vaginal samples against a test population of proteinase K treated semen and vaginal samples. Each blinded sample performed differently with the highest classification

accuracy rate at 100% and the lowest at 28% (Table 4). These large disparities between four different samples not only shows great inter-donor variation amongst vaginal samples but truly shows that a bigger test population sample size may be necessary to gain a full understanding of variability across donor vaginal cell populations. Additionally, this may permit more efficient algorithms for modeling differences in autofluorescence/morphological signature that are driven primarily by cell type rather than donor-to-donor heterogeneity.

Moreover, differences in the number of cell events in each tissue type may affect comparisons, in particular semen and vaginal classifications. The two fertile semen sample types had > 20,000 analyzed cell events whereas vaginal samples were less than 6,000. Higher misclassification of vaginal cells as semen could be due to the fact that the semen cell population is five times larger. In comparison, observing classifications with more balanced sample sizes such as vaginal (5,400) assessed against azoospermic samples (2,800 images) accuracy was overall much better (> 95%). This may be an indication that population disparities might contribute to lesser classification accuracies.

Conclusions

Overall, results from this study suggest that autofluorescence/morphological signatures detected with IFC can be used to differentiate four cell types relevant to sexual assault evidence. Classification rates of individual cell events from each cell population indicate that it may be possible to use signatures to characterize certain components unknown biological samples, in particular those that contain azoospermic and/or degraded semen cell populations when present in a mixture with vaginal and/or rectal cell populations.

However, results also indicated that inter-donor heterogeneity in signatures could have a significant effect on the accuracy of this approach. Assessing not only more overall donors for each cell type but multiple samples from the same donor could work to improve upon inter- and intra-donor variability Testing the validity of the algorithms with blinded mixture samples may provide opportunity for improvement or to solidify the accuracy of results previously seen. Additionally, alternative gating strategies to reduce contributions from non-informative cell events or establishing minimum thresholds for detected cell events of a particular tissue type to conclude that it is present could then be developed to minimize these factors.

Overall, the collection of signatures characterized in this study shows that different cell types from different locations about the body do have naturally occurring morphological and autofluorescent properties that can aid in rapid differentiation. There are currently no robust markers to identify between cell types and even markers for distinguishing between bodily fluids have some overlap in specificity with one another. While there are methods to determine the makeup of samples, nothing is currently in use that does not require further manipulation of the sample, as well. With these methods described above, no hybridization, addition of chemicals, nor DNA extraction are required for semen to be accurately distinguished from vaginal or rectal cell types unlike many of the current methods in casework today.

Looking to the future, utilizing morphological and autofluorescent signatures of distinct cell types to differentiate them could lead to advancements in forensic casework and DNA extraction decisions. Coupling this with a technology such as imaging flow cytometry—that is incredibly rapid (an analysis of a sample in under five minutes) and non-destructive—could allow for a quick survey of a sample to determine cell populations and possible cell type ratios. With this information, not only could the presence of semen be identified faster than a typical microscopic

search for sperm but could be used to direct downstream DNA profiling methods (e.g. Y-STR, mtDNA). Additionally, knowing the cellular makeup of a sexual assault sample can be valuable, probative for the case itself—possibly giving strength or weakening a defendant's claims. Knowing the specific cell types present in a sample is something forensic science has not yet been able to offer, however, information that may be valuable, nonetheless.

References

1. "Sexual Assault Kit Initiative." *FBI*, FBI, 2 Apr. 2019, www.fbi.gov/news/stories/sexual-assault-kit-initiative-040219.

2. "Organizational Priorities & Minimum Topic Requirements for Standards Development." *Organizational Priorities & Minimum Topic Requirements for Standards Development, March 2020*, Forensic Science Standards Board, 3 Mar. 2020, www.nist.gov/system/files/documents/2020/04/23/OSAC%20Organizational%20Priorities_March%202020_CURRENT.pdf.

3. Miller, K. W., Old, J., Fischer, B. R., Schweers, B., Stipinaite, S., and Reich, K. "Developmental Validation of the SPERM HY-LITERTM Kit for the Identification of Human Spermatozoa in Forensic Samples." *Journal of Forensic Sciences, 56*(4), 853-865. doi:10.1111/j.1556-4029.2011.01796.

4. Redhead, P., and Brown, M. K. "The Acid Phosphatase Test Two Minute Cut-off: An Insufficient Time to Detect Some Semen Stains." *Science & Justice*, vol. 53, no. 2, 2013, pp. 187–191., doi:10.1016/j.scijus.2012.09.004.

5. Hellerud, B. B., Bouzga, M., Hoff-Olsen, P., and Mevåg, B. "Semen detection: A retrospective overview from 2010." *Forensic Science International: Genetics Supplement Series, 3*(1). doi:10.1016/j.fsigss.2011.09.057.

6. Stefanidou, M., Alevisopoulos, G., and Spiliopoulou, C. "Fundamental Issues in Forensic Semen Detection." *The West Indian Medical Journal*, vol. 59, no. 3, 2010, p. 280.

7. Romero-Montoya, L., Martinez-Rodriguez, H., Antonio, M., and Argüello-Garcia, R. "Relationship of Spermatoscopy, Prostatic Acid Phosphatase Activity and Prostate-Specific Antigen (p30) Assays with Further DNA Typing in Forensic Samples from Rape Cases." *Forensic Science International*, vol. 206, no. 1-3, 2011, pp. 111–118., doi:10.1016/j.forsciint.2010.07.012.

8. Virkler, K., and Lednev, I. K. "Analysis of Body Fluids for Forensic Purposes: From Laboratory Testing to Non-Destructive Rapid Confirmatory Identification at a Crime Scene." *Forensic Science International*, vol. 188, no. 1, 2009, pp. 1–17., doi:10.1016/j.forsciint.2009.02.013.

9. "TOOLKIT." *SAKI Toolkit: Improving Laboratory Efficiency: A Direct-to-DNA Approach for Testing Sexual Assault Kits*, www.sakitta.org/toolkit/index.cfm?fuseaction=tool&tool=129.

10. Baechtel, F. S. "The Identification and Individualization of Semen Stains." *Forensic Science Handbook*, Prentice Hall, 1988, pp. 347–392.

11. Boward, E. S., and Wilson, S. L. “A Comparison of ABAcard® p30 and RSID™-Semen Test Kits for Forensic Semen Identification.” *Journal of Forensic and Legal Medicine*, vol. 20, no. 8, 2013, pp. 1126–1130., doi:10.1016/j.jflm.2013.09.007.

12. Roewer, L. “Y Chromosome STR Typing in Crime Casework.” *Forensic Science, Medicine, and Pathology*, vol. 5, no. 2, 2009, pp. 77–84., doi:10.1007/s12024-009-9089-5.

13. Soares-Vieira, J. A., Correia Billerbeck, A., Miazato, I., Sadayo, E., Zampieri, R. A., Figaro Gattas, G. J., et al. “Y-STRs in Forensic Medicine: DNA Analysis in Semen Samples of Azoospermic Individuals.” *Journal of Forensic Sciences*, vol. 52, no. 3, 2007, pp. 664–670., doi:10.1111/j.1556-4029.2007.00433.x.

14. Cotton, R. W., and Fisher, M. B. “Review: Properties of Sperm and Seminal Fluid, Informed by Research on Reproduction and Contraception.” *Forensic Science International: Genetics*, vol. 18, 2015, pp. 66–77., doi:10.1016/j.fsigen.2015.03.009.

15. “National Best Practices for Sexual Assault Kits: A Multidisciplinary Approach.” *National Institute of Justice*, nij.ojp.gov/topics/articles/national-best-practices-sexual-assault-kits-multidisciplinary-approach.

16. “Recommendations for the Efficient DNA Processing of Sexual Assault Evidence Kits.” *Swgdam*, 5 Dec. 2016, www.swgdam.org/publications.

17. Visser, M., Zubakov, D., Ballantyne, K., and Kayser, M. “MRNA-Based Skin Identification for Forensic Applications.” *International Journal of Legal Medicine*, vol. 125, no. 2, 2011, pp. 253–263., doi:10.1007/s00414-010-0545-2.

18. Sijen, T. "Molecular approaches for forensic cell type identification: on mRNA, miRNA, DNA methylation and microbial markers." *Forensic Science International: Genetics* 18 (2015): 21-32.

19. Hanson, E., Ingold, S., Haas, C., and Ballantyne, J. "Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing." *Forensic Science International: Genetics* 34 (2018): 206-221.

20. Fleming, R., Harbison, S., and Lin, M. “New RNA Methods for the Identification of Body Fluids and Cell Types.” *Forensic Science International. Genetics Supplement Series*, vol. 4, no. 1, 2013, pp. e87–e88., doi:10.1016/j.fsigss.2013.10.044.

21. Harbison, S., and Fleming, R. "Forensic body fluid identification: state of the art." *Res Rep Forensic Med Sci* 6.2016 (2016): 11-23.

22. Sauer, E., Exta, A., Cachee, P., and Courts, C. "Identification of organ tissue types and skin from forensic samples by microRNA expression analysis." *Forensic Science International: Genetics* 28 (2017): 99-110.

23. Silva, D. S.B.S., Antunes, J., Balamurugan, K., Duncan, G., Alho C. S., and McCord, B. "Developmental Validation Studies of Epigenetic DNA Methylation Markers for the Detection of Blood, Semen and Saliva Samples." *Forensic Science International: Genetics*, vol. 23, 2016, pp. 55–63., doi:10.1016/j.fsigen.2016.01.017.

24. Lindenbergh, A., Maaskant, P., and Sijen, T. "Implementation of RNA Profiling in Forensic Casework." *Forensic Science International: Genetics*, vol. 7, no. 1, 2013, pp. 159–166., doi:10.1016/j.fsigen.2012.09.003.

25. Tian, H., Lv, M., Li, A., Peng, D., Tan, Y., Wang, H., et al. "Semen-specific miRNAs: Suitable for the distinction of infertile semen in the body fluid identification?." *Forensic Science International: Genetics* 33 (2018): 161-167.

26. Frumkin, D., Wasserstrom, A., Budowle, B., and Davidson, A. "DNA Methylation-Based Forensic Tissue Identification." *Forensic Science International: Genetics*, vol. 5, no. 5, 2011, pp. 517–524., doi:10.1016/j.fsigen.2010.12.001.

27. Baron, U., Turbachova, I., Hellwag, A., Eckhardt, F., Berlin, K., Hoffmüller, U., Gardina, P., and Olek, S. "DNA Methylation Analysis as a Tool for Cell Typing." *Epigenetics*, vol. 1, no. 1, 2014, pp. 56–61., doi:10.4161/epi.1.1.2643.

28. Antunes, J., Silva, D. S.B. S., Balamurugan, K., Duncan, G., Alho C. S., and McCord, B. "Forensic Discrimination of Vaginal Epithelia by DNA Methylation Analysis through Pyrosequencing." *Electrophoresis*, vol. 37, no. 21, 2016, pp. 2751–2758., doi:10.1002/elps.201600037.

29. Eipel, M., Mayer, F., Arent, T., Marcelo, P., Birkhofer, C., Gerstenmaier, U., Costa, I., Ritz-Timme, S., and Wagner, W. "Epigenetic Age Predictions Based on Buccal Swabs Are More Precise in Combination with Cell Type-Specific DNA Methylation Signatures." *Aging*, vol. 8, no. 5, 2016, pp. 1034–1048., doi:10.18632/aging.100972.

30. Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V., and Morrissey, P. "Cellular Image Analysis and Imaging by Flow Cytometry." *Clinics in Laboratory Medicine*, vol. 27, no. 3, 2007, pp. 653–670., doi:10.1016/j.cll.2007.05.008.

31. Barteneva, N. S., and Vorobjev, I. "Imaging Flow Cytometry." *Journal of Histochemistry & Cytochemistry*, vol. 60, no. 10, 2012, pp. 723–733., doi:10.1369/0022155412453052.

32. Brocato, E. R., Philpott, M. K., Connon, C. C., and Ehrhardt, C. J. "Rapid Differentiation of Epithelial Cell Types in Aged Biological Samples Using Autofluorescence and Morphological Signatures." *Plos One*, vol. 13, no. 5, 2018, doi:10.1371/journal.pone.0197701.

33. Sijen, T. "Molecular Approaches for Forensic Cell Type Identification: On MRNA, MiRNA, DNA Methylation and Microbial Markers." *Forensic Science International: Genetics*, vol. 18, 2015, pp. 21–32., doi:10.1016/j.fsigen.2014.11.015.

34. “4.2 Epithelial Tissue.” *Anatomy & Physiology*, by J. Gordon Betts, OpenStax College, Rice University, 2013.

35. Dettmeyer, R. B., author. *Forensic Histopathology Fundamentals and Perspectives*. Cham: Springer International Publishing: Imprint: Springer, 2018.

36. Tanaka, E., Noguchi, T., Nagai, K., Akashi, Y., Kawahara, K., and Shimada, T. “Morphology of the epithelium of the lower rectum and the anal canal in the adult human.” *Medical Molecular Morphology*, *45*(2), 72-79. doi:10.1007/s00795-011-0541-8.

37. Randall, B. “Glycogenated Squamous Epithelial Cells as a Marker of Foreign Body Penetration in Sexual Assault.” *Journal of Forensic Sciences*, vol. 33, no. 2, 1988, pp. 511–514., doi:10.1520/JFS11965J.

38. Randall, B, and Riis, R. E. “Penile glycogenated epithelial cells as an indicator of recent vaginal intercourse.” *American journal of clinical pathology* vol. 84,4 (1985): 524-9. doi:10.1093/ajcp/84.4.524

39. Cortijo, M., Steinberg, I. Z., and Shaltiel, S. “Fluorescence of Glycogen Phosphorylase b.” *Journal of Biological Chemistry*, vol. 246, no. 4, 1971, p. 933.

40. Graham, E. K., Loten, M., Thompson, J. M., Drobac, J., and Gopalakrishnan, A. “Developmental Validation of the Casework Direct Kit, Custom: A Method for the Rapid Processing of Casework Samples.” *Promega Corporation*, 2018, www.promega.com/products/forensic-dna-analysis-ce/preprocessing-and-differential-extraction/casework-direct-kit/?catNum=AX4560.

41. Palermo, G. D., Neri, Q. V., Cozzubbo, T., Cheung, S., Pereira, N., and Rosenwaks, Z. “Shedding Light on the Nature of Seminal Round Cells.” *PLoS ONE*, vol. 11, no. 3, 2016, p. e0151640., doi:10.1371/journal.pone.0151640.

42. Bellastella, G., Cooper, T. G., Battaglia, M., Strose, A., Torres, I., Hellenkemper, B., Soler, C., and Sinisi, A. A. “Dimensions of human ejaculated spermatozoa in Papanicolaou-stained seminal and swim-up smears obtained from the Integrated Semen Analysis System (ISAS(®)).” *Asian journal of andrology* vol. 12,6 (2010): 871-9. doi:10.1038/aja.2010.90.

43. Wong, H., and Mihalovich J. “Automation of the Differential Digestion Process of Sexual Assault Evidence.” *Journal of Forensic Sciences*, vol. 64, no. 2, 2019, pp. 539–550., doi:10.1111/1556-4029.13877.

44. Chapman, B. R. B., Blackwell, S. J., and Müller, L. H. “Forensic Techniques for the Isolation of Spermatozoa from Sexual Assault Samples - A Review.” *Forensic Science Review*, vol. 32, no. 2, 2020, pp. 105–116.

45. Amano, T., Kunimi, K., and Ohkawa, M. “Fluorescence Spectra from Human Semen and Their Relationship with Sperm Parameters.” *Archives of Andrology, 36*(1), 9-15. doi:10.3109/01485019608987879.

46. Evers, H., Birngruber, C. G., Ramsthaler, F., Müller, U., Brück, S., and Verhoff, M. A. “Unterscheidung von Epithelzell-Typen anhand des Zelldurchmessers” [Differentiation of epithelial cell types by cell diameter]. *Archiv fur Kriminologie* vol. 228,1-2 (2011): 11-9.

47. Paterson, S. K., Jensen, C. G., Vintiner, S. K., and McGlashan, S. R. “Immunohistochemical Staining as a Potential Method for the Identification of Vaginal Epithelial Cells in Forensic Casework.” *Journal of Forensic Sciences*, vol. 51, no. 5, 2006, pp. 1138–1143., doi:10.1111/j.1556-4029.2006.00234.x.

Appendix.

Table 1. Total number of analyzable captured images per sample that fell within the specified gate. Noticeable differences are present between the number of vaginal cells captured to the number of rectal and penile cells. This is no longer of issue, however, when these same samples were assessed with the large cell gate. The same can be noted with azoospermic semen compared to the other semen sample types.

Sample Type:	Small Cell Gate	Large Cell Gate
Vaginal Samples	**5462**	**2715**
Rectal Samples	**13085**	**1109**
Penile Samples	**24169**	**1754**
Semen	**20090**	**-**
Proteinase K Semen	**19108**	**-**
Azoospermic Semen	**2841**	**-**

Table 2. Complete list of feature values obtained for each single-cell image captured and what their correlation to what they measure per image. Bolded feature values are those utilized when running the initial DFA analysis of samples.

Feature Values:					
Size	**Location**	**Shape**	**Texture**	**Signal Strength**	**Other**
Area	**Angle**	**Aspect Ratio**	**Bright Detail Intensity R3 and R7**	**Intensity**	**H. Contrast Mean and Standard Deviation**
Diameter	Max Contour	**Aspect Ratio Intensity**	**Contrast**	**Max Pixel**	**H. Correlation Mean and Standard Deviation**
Height	Spot Distance	**Circularity**	Gradient RMS	**Mean Pixel**	**H. Energy Mean and Standard Deviation**
Length		**Elongatedness**	Gradient Max	**Median Pixel**	**H. Entropy Mean and Standard Deviation**
Width		Lobe Count	Spot Count	**Min Pixel**	**H. Homogeneity Mean and Standard Deviation**
Thickness Min/Max		Shape Ratio		**Spot Intensity Min and Max**	**H. Variance Mean and Standard Deviation**
Major and Minor Axis		Symmetry 2, 3, 4			Internalization
Major and Minor Intensity					Objects per mL
Perimeter					
Spot Area Min					

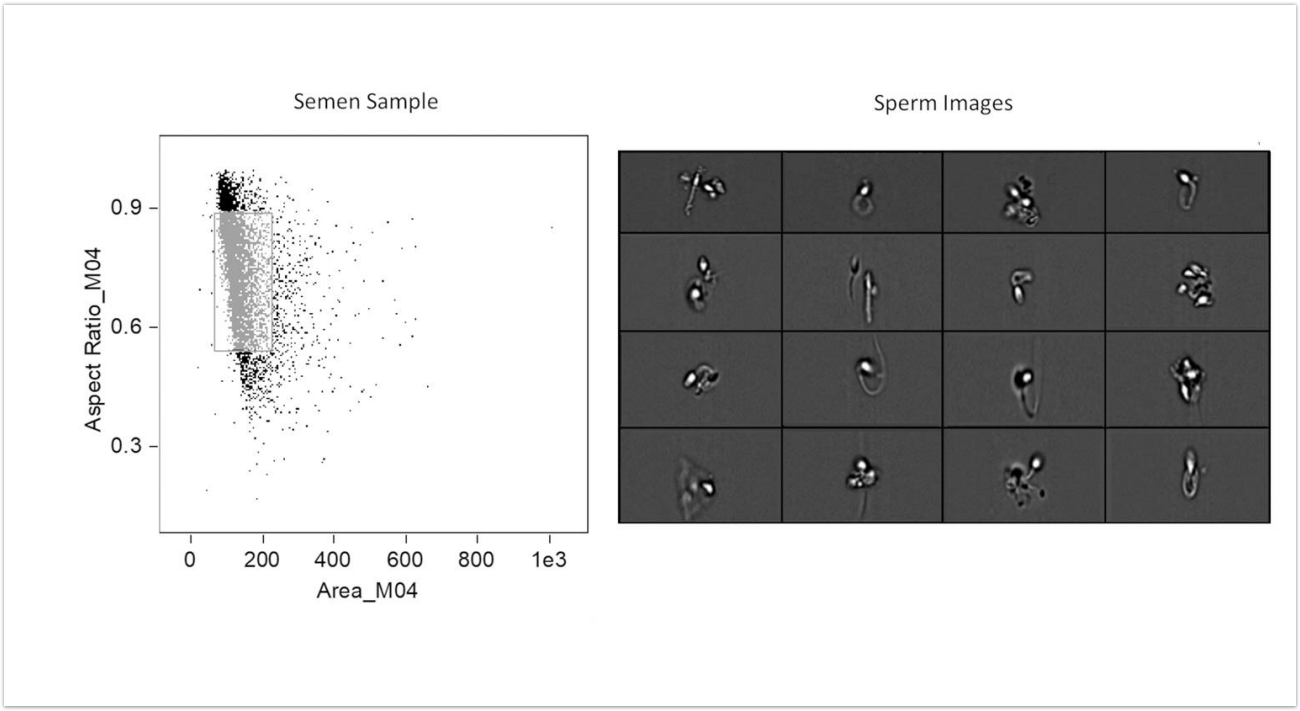
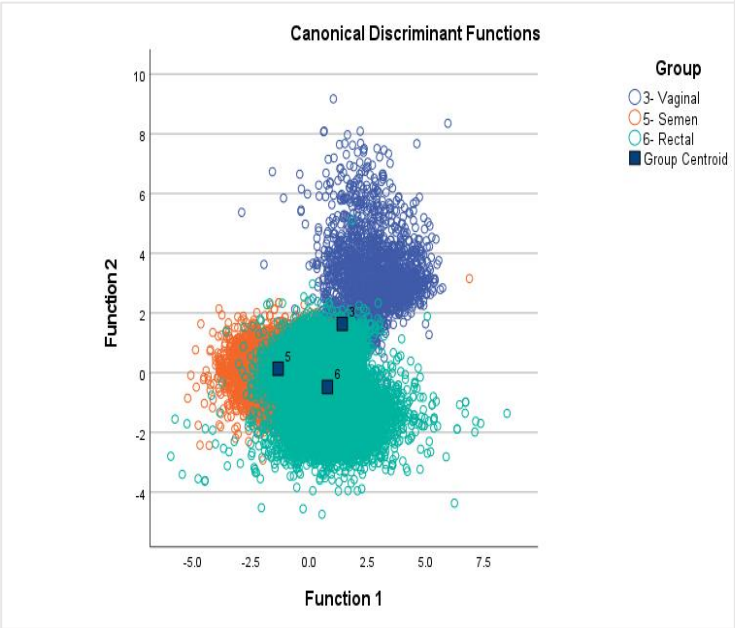


Figure 1. Semen sample plot analyzed with the sperm cell gate (seen in grey) to obtain the sperm-only fraction of the sample. Sperm gate utilized for all semen samples. Sperm images in the Brightfield channel (Channel 04) are shown in the right panel.



Classification Results[a,c]

		Predicted Group Membership				
		Group	3	5	6	Total
Original	Count	3	4141	955	366	5462
		5	292	18966	832	20090
		6	2984	3843	17342	24169
	%	3	75.8	17.5	6.7	100.0
		5	1.5	94.4	4.1	100.0
		6	12.3	15.9	71.8	100.0
Cross-validated[b]	Count	3	4133	957	372	5462
		5	294	18954	842	20090
		6	3005	3874	17290	24169
	%	3	75.7	17.5	6.8	100.0
		5	1.5	94.3	4.2	100.0
		6	12.4	16.0	71.5	100.0

a. 81.4% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 81.2% of cross-validated grouped cases correctly classified.

Figure 2. Discriminant Function Analysis of semen, vaginal, and rectal cells (small cell gate) and the accompanying classification results. Overlap for all three sample populations is clear, especially rectal event overlap with semen and vaginal samples. Classification accuracy for vaginal and rectal samples was observed below 80%, while semen was classified with an accuracy rate of 94%. Vaginal and rectal samples misclassified as semen with an accuracy rate around 16-17% each.

Classification Results[a,c]

		Group	Predicted Group Membership 3	5	Total
Original	Count	3	4582	880	5462
		5	438	19652	20090
	%	3	83.9	16.1	100.0
		5	2.2	97.8	100.0
Cross-validated[b]	Count	3	4576	886	5462
		5	462	19628	20090
	%	3	83.8	16.2	100.0
		5	2.3	97.7	100.0

a. 94.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 94.7% of cross-validated grouped cases correctly classified.

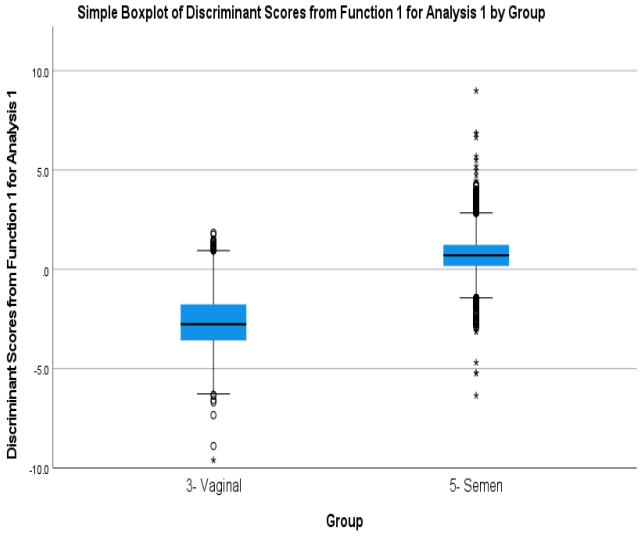


Figure 3. Classification results for semen (5) and vaginal cell populations (3) and boxplot of subsequent discriminant functional analysis. Semen retains a high accuracy rate of classification at around 97%, however, vaginal cells misclassify as semen at a rate of 16.2%.

Classification Results[a,c]

		Group	Predicted Group Membership 5	6	Total
Original	Count	5	19240	850	20090
		6	3643	13500	17143
	%	5	95.8	4.2	100.0
		6	21.3	78.7	100.0
Cross-validated[b]	Count	5	19221	869	20090
		6	3660	13483	17143
	%	5	95.7	4.3	100.0
		6	21.3	78.7	100.0

a. 87.9% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 87.8% of cross-validated grouped cases correctly classified.

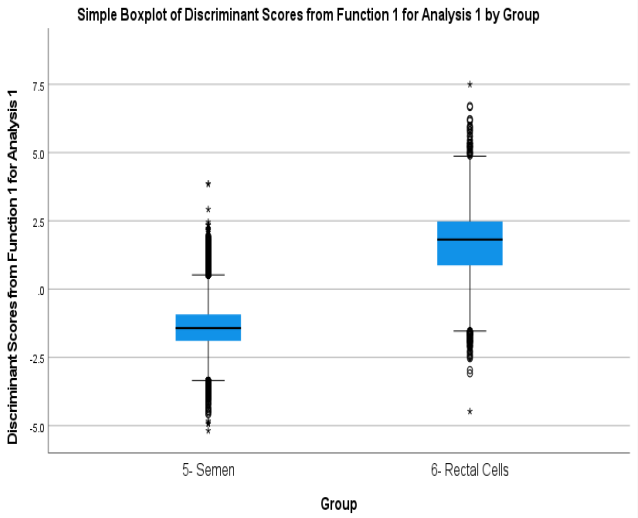
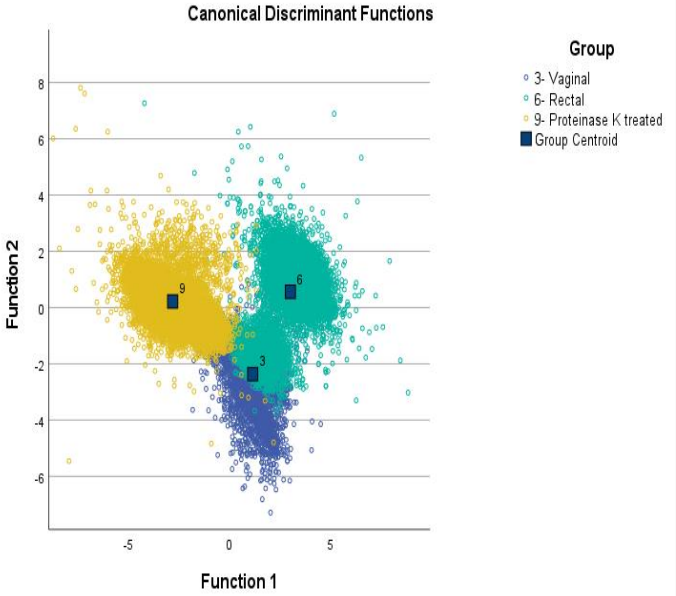


Figure 4. Rectal cells (6) and semen (5) classification results and corresponding boxplot of the discriminant analysis. Similar to vaginal cell classification, rectal cells misclassify as semen at a rate of 21.3%.



Classification Results[a,c]

			Predicted Group Membership			
		Group	3	6	9	Total
Original	Count	3	4377	954	131	5462
		6	7292	16841	36	24169
		9	505	63	18540	19108
	%	3	80.1	17.5	2.4	100.0
		6	30.2	69.7	.1	100.0
		9	2.6	.3	97.0	100.0
Cross-validated[b]	Count	3	4364	965	133	5462
		6	7316	16814	39	24169
		9	510	64	18534	19108
	%	3	79.9	17.7	2.4	100.0
		6	30.3	69.6	.2	100.0
		9	2.7	.3	97.0	100.0

a. 81.6% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 81.5% of cross-validated grouped cases correctly classified.

Figure 5. Discriminant Function Analysis of proteinase K treated semen, vaginal, and rectal cells (small cell gate) and the accompanying classification results. Clear population clusters were noted for all three sample populations, however, proteinase K treated samples had the clearest separation. Some event overlap noted between rectal and vaginal samples. Classification accuracy for vaginal and rectal cell populations was observed at less than 80%, while semen samples classified with an accuracy of 97%.

Classification Results[a,c]

		Group	Predicted Group Membership 3	9	Total
Original	Count	3	5413	49	5462
		9	107	19001	19108
	%	3	99.1	.9	100.0
		9	.6	99.4	100.0
Cross-validated[b]	Count	3	5408	54	5462
		9	122	18986	19108
	%	3	99.0	1.0	100.0
		9	.6	99.4	100.0

a. 99.4% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 99.3% of cross-validated grouped cases correctly classified.

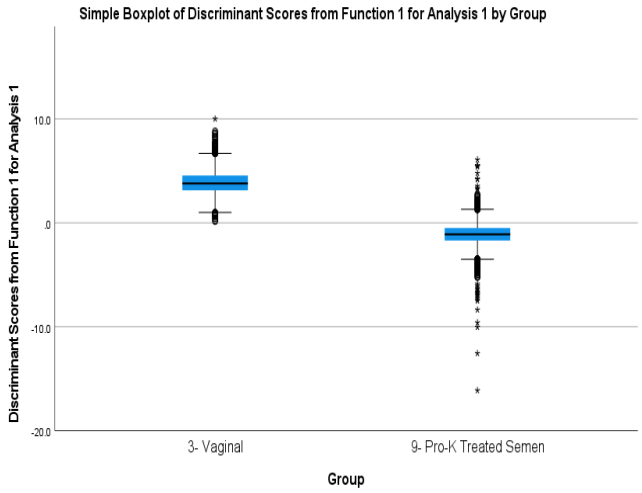


Figure 6. Classification results for proteinase K treated (9) and vaginal samples (3). The accuracy (at or above 99%) indicates a clear differentiation between sample types, possibly due to decreased fluorescence of the proteinase K samples.

Classification Results[a,c]

		Group	Predicted Group Membership 6	9	Total
Original	Count	6	15634	35	15669
		9	43	19065	19108
	%	6	99.8	.2	100.0
		9	.2	99.8	100.0
Cross-validated[b]	Count	6	15628	41	15669
		9	49	19059	19108
	%	6	99.7	.3	100.0
		9	.3	99.7	100.0

a. 99.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 99.7% of cross-validated grouped cases correctly classified.

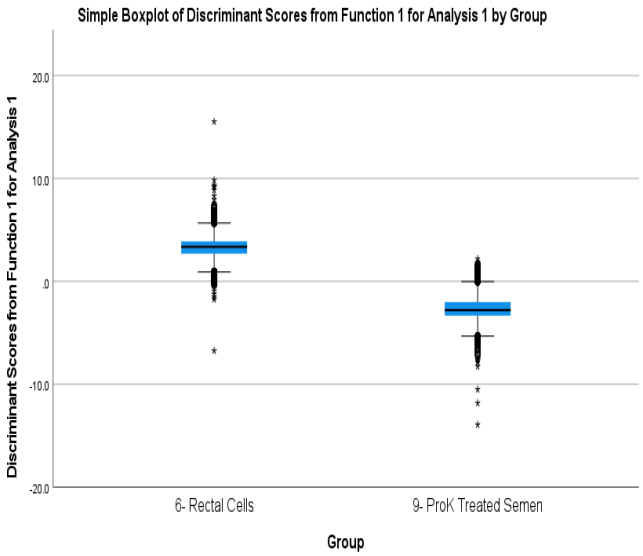
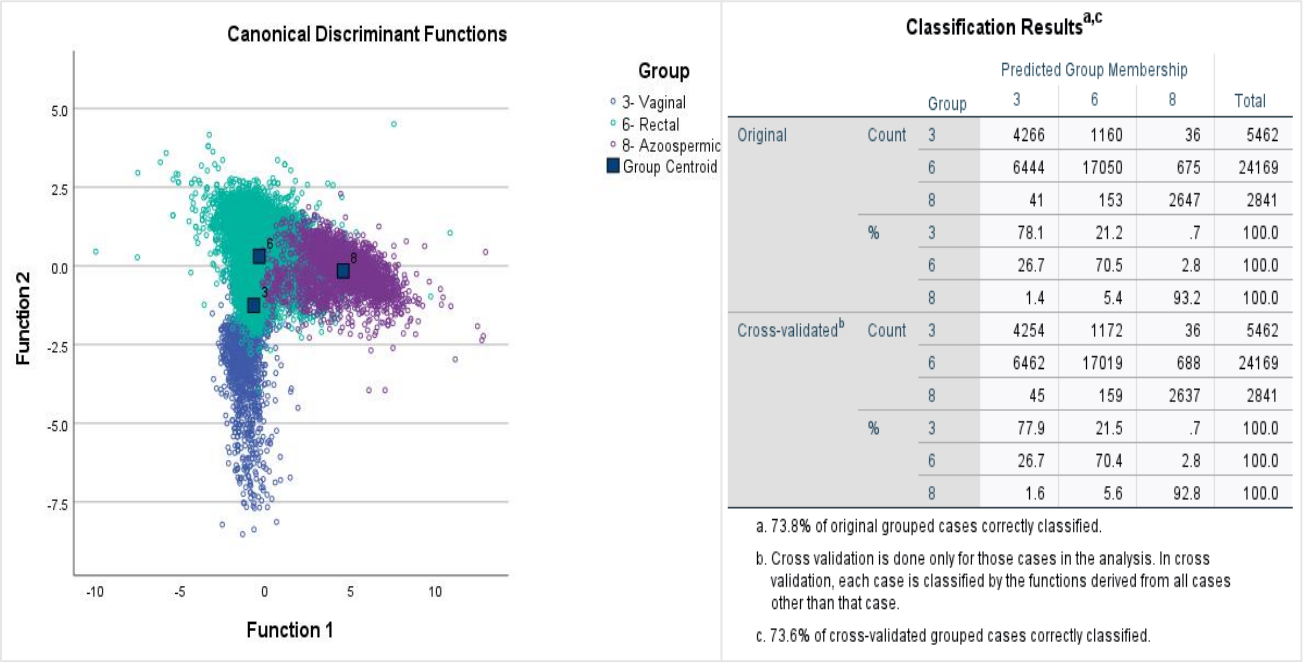


Figure 7. Rectal cells (6) and proteinase K treated semen (9) classification results and corresponding boxplot of the discriminant analysis. Accuracy rates for both cell populations are near 100% accuracy.



Classification Results[a,c]

			Predicted Group Membership			
		Group	3	6	8	Total
Original	Count	3	4266	1160	36	5462
		6	6444	17050	675	24169
		8	41	153	2647	2841
	%	3	78.1	21.2	.7	100.0
		6	26.7	70.5	2.8	100.0
		8	1.4	5.4	93.2	100.0
Cross-validated[b]	Count	3	4254	1172	36	5462
		6	6462	17019	688	24169
		8	45	159	2637	2841
	%	3	77.9	21.5	.7	100.0
		6	26.7	70.4	2.8	100.0
		8	1.6	5.6	92.8	100.0

a. 73.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 73.6% of cross-validated grouped cases correctly classified.

Figure 8. Discriminant Function Analysis of azoospermic seminal fluid, vaginal, and rectal cells (small cell gate) and the accompanying classification results. Separation of the azoospermic group centroid and cluster noted with overlap noted for all three groups. Classification accuracy was poor between rectal and vaginal cells while azoospermic samples were classified correctly at an accuracy of 92%.

Classification Results[a,c]

		Group	Predicted Group Membership 3	8	Total
Original	Count	3	5376	86	5462
		8	103	2738	2841
	%	3	98.4	1.6	100.0
		8	3.6	96.4	100.0
Cross-validated[b]	Count	3	5366	96	5462
		8	111	2730	2841
	%	3	98.2	1.8	100.0
		8	3.9	96.1	100.0

a. 97.7% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 97.5% of cross-validated grouped cases correctly classified.

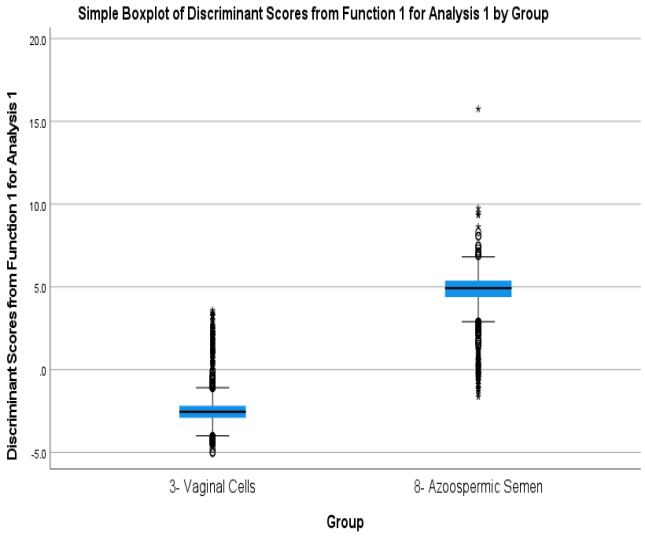


Figure 9. Vaginal cell (3) and azoospermic semen (8) classification results and corresponding boxplot of the discriminant analysis. Both samples classify at a rate above 96% with azoospermic semen samples misclassifying most at around 4%.

Classification Results[a,c]

		Group	Predicted Group Membership 6	8	Total
Original	Count	6	23837	332	24169
		8	100	2741	2841
	%	6	98.6	1.4	100.0
		8	3.5	96.5	100.0
Cross-validated[b]	Count	6	23815	354	24169
		8	101	2740	2841
	%	6	98.5	1.5	100.0
		8	3.6	96.4	100.0

a. 98.4% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 98.3% of cross-validated grouped cases correctly classified.

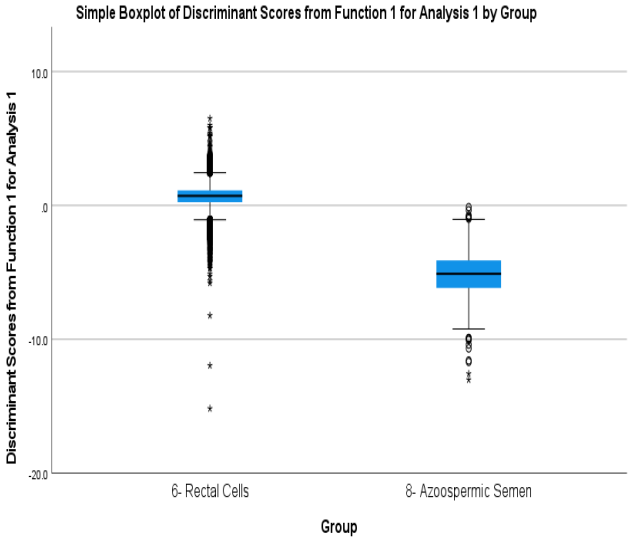


Figure 10. Rectal cell (6) and azoospermic semen (8) classification results and corresponding boxplot of the discriminant analysis. Again, both samples are classifying at a rate above 96% with azoospermic semen samples again misclassifying around 4%.

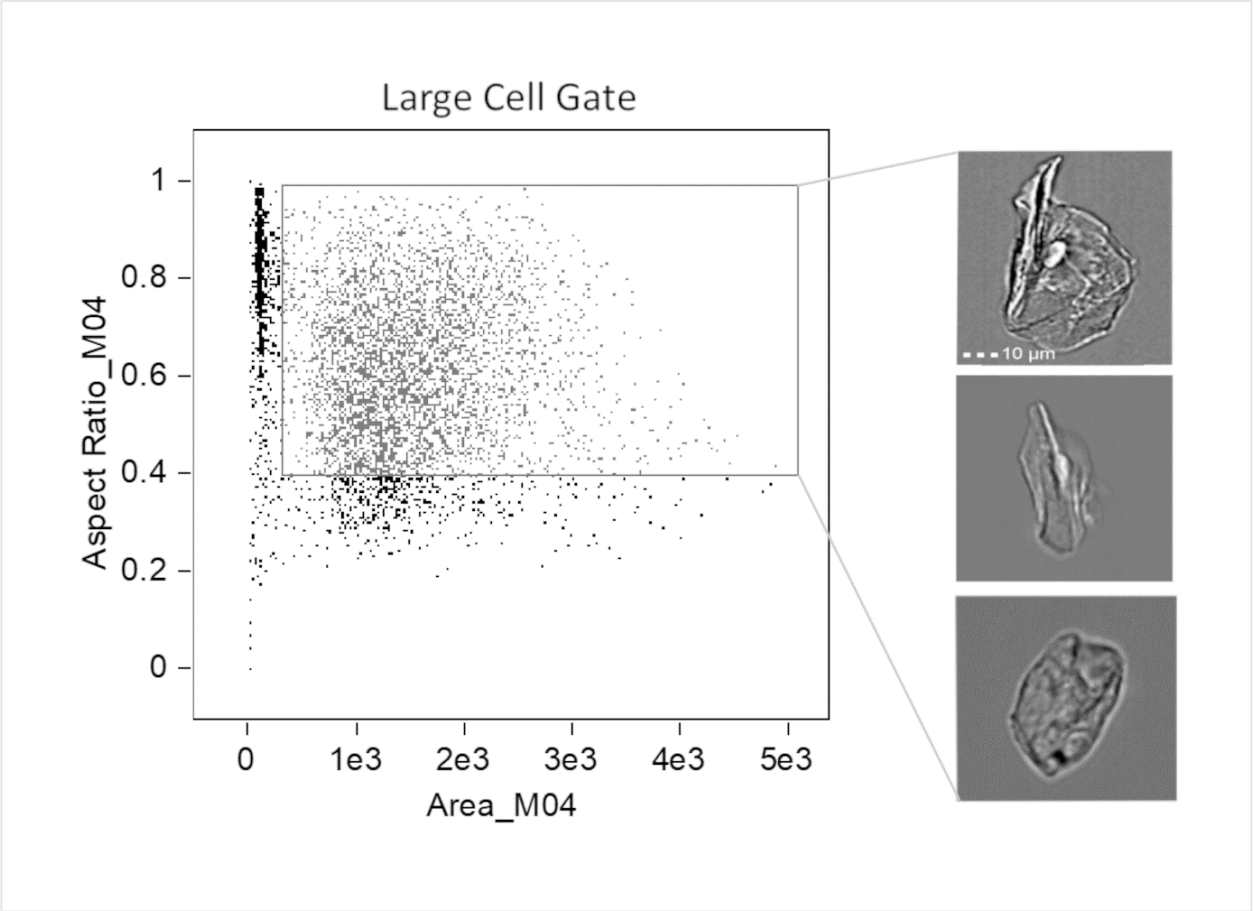
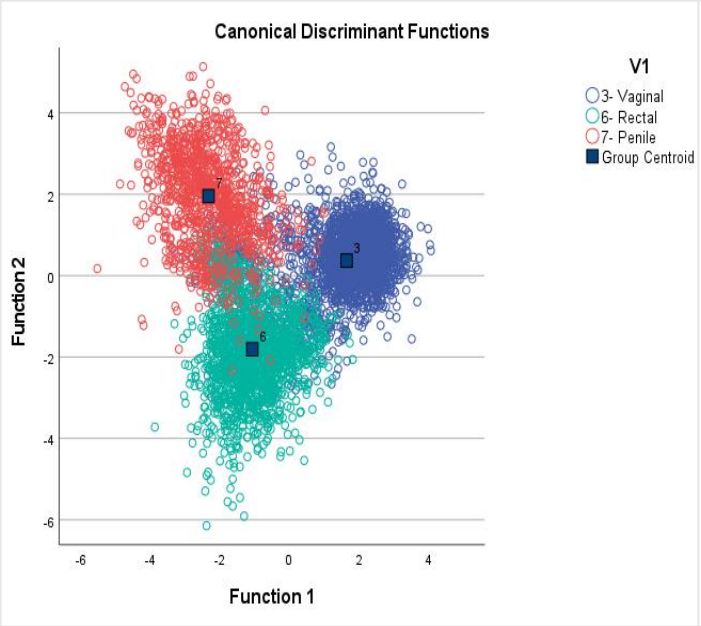


Figure 11. Large cell gate example representing where the majority of images for large cell populations (top to bottom: vaginal, rectal, penile cells) are typically found in a sample plot.

Figure 12. Image gallery for three common epithelial types in sexual assault samples. IFC Brightfield images (Chanel 04) for vaginal cells (columns 1-3), rectal cells (columns 4-6), and penile cells (columns 7-9).



Classification Results[a,c]

			Predicted Group Membership			
		Group	3	6	7	Total
Original	Count	3	5227	240	171	5638
		6	91	1475	188	1754
		7	42	47	1020	1109
	%	3	92.7	4.3	3.0	100.0
		6	5.2	84.1	10.7	100.0
		7	3.8	4.2	92.0	100.0
Cross-validated[b]	Count	3	5212	250	176	5638
		6	101	1458	195	1754
		7	44	52	1013	1109
	%	3	92.4	4.4	3.1	100.0
		6	5.8	83.1	11.1	100.0
		7	4.0	4.7	91.3	100.0

a. 90.8% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 90.4% of cross-validated grouped cases correctly classified.

Figure 13. Discriminant Function Analysis of vaginal, penile and rectal cells (large gate) and accompanying classification results. Clear separation of population clusters and centroids for all three cell samples noted. Vaginal and penile samples classifying at a rate above 90% while rectal samples accurate classify around 83%.

Classification Results[a,c]

		Group	Predicted Group Membership 6	7	Total
Original	Count	6	1681	73	1754
		7	55	1054	1109
	%	6	95.8	4.2	100.0
		7	5.0	95.0	100.0
Cross-validated[b]	Count	6	1666	88	1754
		7	67	1042	1109
	%	6	95.0	5.0	100.0
		7	6.0	94.0	100.0

a. 95.5% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 94.6% of cross-validated grouped cases correctly classified.

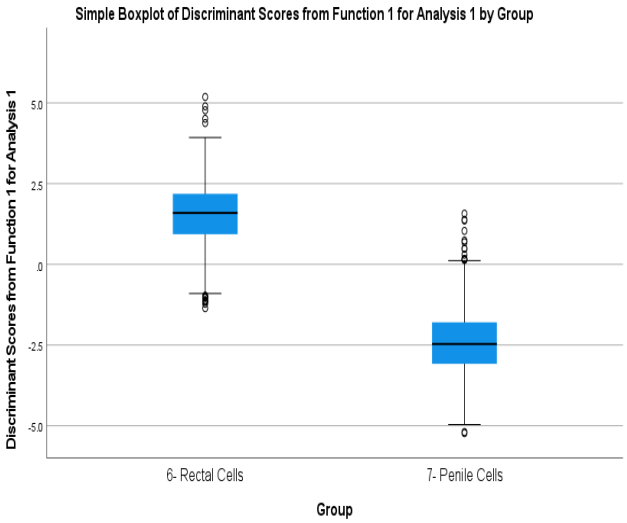


Figure 14. Rectal (6) and penile cell (7) classification results and corresponding DFA boxplot. Rectal and penile samples classifying at 94% and 95% respectively.

Classification Results[a,c]

		Group	Predicted Group Membership 3	7	Total
Original	Count	3	5444	194	5638
		7	57	1052	1109
	%	3	96.6	3.4	100.0
		7	5.1	94.9	100.0
Cross-validated[b]	Count	3	5431	207	5638
		7	67	1042	1109
	%	3	96.3	3.7	100.0
		7	6.0	94.0	100.0

a. 96.3% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 95.9% of cross-validated grouped cases correctly classified.

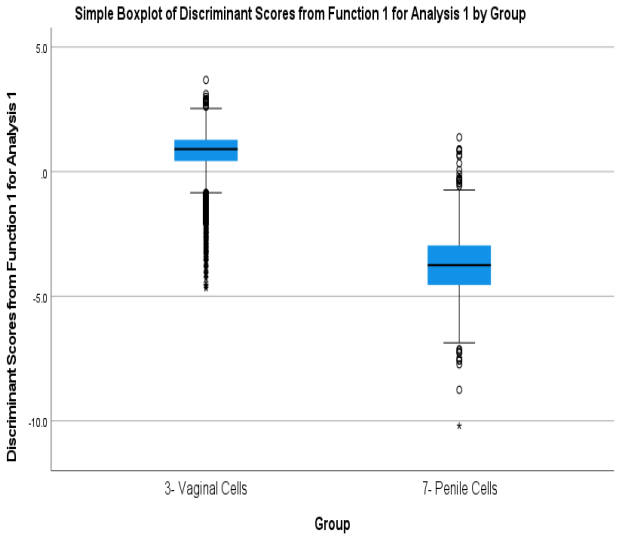


Figure 15. Vaginal (3) and penile cell (7) sample classification results and corresponding DFA boxplot. Vaginal samples accurately classifying at a rate around 96% and penile cells at a rate of 94%.

Table 3. Blinded classification results for withheld donor populations of proteinase K semen samples. All three samples have high and similar classification rates indicating homogeneity amongst semen samples.

		Predicted Cell Type		
Cell Type	**ID**	**Semen**	**Vaginal**	**Classification %**
Pro K Semen	#3	6085	90	98.5
Pro K Semen	#4	6443	17	99.7
Pro K Semen	#9	6458	15	99.2

Table 4. Blinded classification results for withheld donor populations of vaginal samples. Poor accuracy rates and inconsistencies indicate high heterogeneity amongst vaginal donor cell populations.

		Predicted Cell Type		
Cell Type	**ID**	**Proteinase K Treated Semen**	**Vaginal**	**Classification %**
Vaginal	1022	0	866	100
Vaginal	1028	193	75	28.0
Vaginal	2368	156	218	58.3
Vaginal	4017	151	370	71.0

Vita

Taylor Renee Moldenhauer was born April 6, 1995 in Lawrence, Kansas. She attended Lawrence High School as well as the University of Kansas. It was at KU where she served as a board member for the American Microbiology Society KU Chapter and the KU Equestrian team. Additionally, she contributed to the lab of Dr. Stuart Macdonald and her research endeavors in this lab led to a co-authorship on the paper "Naturally Segregating Variation at *Ugt86Dd* Contributes to Nicotine Resistance in *Drosophila melanogaster"*. In 2017 she graduated with a Bachelor of Science in Microbiology. After a year of working in a forensically accredited drug-testing lab, Taylor decided to pursue a career in forensic science. In 2019 she began attending Virginia Commonwealth University Master of Science in Forensic Science program as a biology track student. She was a Graduate Teaching Assistant for the first year but has now taken on a position as a Graduate Research Assistant for the lab of Dr. Christopher Ehrhardt. She aims to present her research work at AAFS in February 2021.